Review of

Evaluation of ARM Tethered Balloon System instrumentation for supercooled liquid water and distributed temperature sensing in mixed-phase Arctic clouds

by Darielle Dexheimer et al.

General Comments:

The manuscript describes an evaluation of tethered aerostats helikites instrumented for cloud measurements. This reviewer has several problems with the way this manuscript is presented, and how the measurements are evaluated.

1. The manuscript approaches the subject as if its TBS is a novel sampling technique that has never before been attempted, when in fact, the literature contains references to cloud sampling with tethered balloons for nearly the past 40 years. Kitchen and McClatchey (1981) flew an instrumented aerostat in cumulus clouds in the UK. Several investigators have followed suit, including measurements in the Arctic (Lawson et al. 2011) and at the South Pole (Lawson and Gettelman 2014). The manuscript needs to provide context for their research by discussing the previous efforts and explain why their current approach is an improvement, or how it differs.

2. The methodology used to process and present the measurements does not do justice to the larger variance in the different techniques being compared. While the variance can be seen in some of the measurements (e.g. time series in Fig. 3; scatterplots in Fig. 7), the overall message presented is that there is relatively good agreement in the measurements and that therefore they are useful. In fact, it is not at all clear that the vibrating wire technique is of any value in quantitatively measuring SLWC. The manuscript compares mean values of SLWC measured from the two vibrating wire techniques to validate the instruments when the variance in the measurements is larger than the mean; this is not an acceptable method for validating instrument performance. I agree that the technique may be useful as an SLWC detection method, and perhaps it can differentiate between low values i.e., < a few tenths g m^{-3} and high values (> ~ 1 g m^{-3}), but even this cannot be validated since high SLWC measurements are not presented.

3. I do feel that the DTS measurements are sufficiently accurate to be useful, but the validation is against a radiosonde. If the comparison is good, why not just use radiosonde data? Certainly it is more cost effective than the complexity involved with launching a tethered aerostat or helikite. One possible use of the DTS measurements is to better define low-level inversions, but I did not find a meteorological use for this presented in the paper.

4. Before the manuscript is fit for publication it should be modified considerably. The measurements are presented in a manner that embellishes their actual efficacy. The manuscript should present the data in an biased manner and focus on the ability of the TBS to make routine, long-term measurements, but eliminate the impression that the instruments themselves are useful for making quantitative measurements.

Specific Comments:

Introduction: Add TBS background references and discuss.

Page 3, Lines 11 – 14: This is incorrect. There are much larger helikytes currently in service. For example, Bodenschatz, in Goettingen has deployed a 250 m^3 helikyte from a ship in the Arctic.

Figure 3: The figure and text states that the difference in mean values of SLWC between the two Anasphere devices is 0.01, 0.02 and 0.06 g m^{-3}.   This is meaningless without also giving some measurement of the variance, standard deviation or show a scatterplot. Even two noise signals can have the same mean. The figure suggests that there is little correlation between the two devices.

Figure 7: I applaud the authors for showing the scatterplots in this figure, but the interpretation of the results does not fit the data. I am not sure how an R^2 value is even relevant. The scatterplot and time series tell the story. There is virtually no correlation in the measurements. The best that one can ascertain here is that the TBS is an SLWC detector.

Figures 8 and 9: This is a good representation of the measurements, but again, the interpretation is off the mark. The statement on p. 16,

"Overall, the two sensors seem to provide broadly similar SLWC profiles, but not without some discrepancies. This increases confidence in the use of each of them and provides independent verification of the measurements."

grossly overstates the efficacy of the measurements. The profiles in Figure 9 often diverge and differ by factors of 2 to 3. The authors need to show measurements with better correlation over a much larger range of SLWC's, and the sensors need to agree with SLWCad to convince readers that the instruments can be used quantitatively. Several aircraft campaigns have shown that single-layer ice-free Arctic stratus clouds typically have adiabatic SLWC profiles up to 60 to 90% of the distance between cloud base and cloud top. Of course, this will be challenging if the balloons cannot achieve higher altitudes.

Page 17, Lines 4 – 5: Equation 3 computes the adiabatic maximum SLWC. Since the vibrating wires measured a higher value than adiabatic, the measurements exceed the theoretical maximum. Also, once again, only reporting means does not show the variance in the measurements, which should also be reported, preferably by showing a scatterplot.

Page 17, Line 8: I have examined the Sand et al. (1984) article and I cannot find any text or figures that support the claim in the manuscript. Please show how 36% of the samples < 0.05 g m^{-3} was derived from the Sand paper. Also, the minimum measurement sensitivity of the J-W used in the Sand study was 0.1 g m^{-3} due to drift and noise affecting the baseline. The FSSP was used for lower values, but its fundamental measurement is drop size and raising drop size to the third power increases sizing errors commensurately. Thus, the FSSP is not a fundamental measurement of LWC. Reference to the Sand paper is a stretch and appears to be an unsubstantiated attempt to validate the vibrating wire measurements.

Conclusions: The conclusions are very misleading. The measurements are limited to very low values of SLWC, so the dynamic range over which the measurements are constrained is tiny. Means are compared within this very limited dynamic range, making the comparison look good (e.g., the means of the measurements compared within 0.1 g m^{-3}). This a totally unacceptable methodology for presenting the measurements. Showing the time series comparisons and scatterplots is representative and these figures are the saving grace of this manuscript. The paper should focus on the mechanics of how the TBS systems are operated, and that they can be successfully operated for extended periods of time. The vibrating wire technology is used on commercial aircraft to detect icing conditions, not to quantify icing rate, even though there was an attempt in the 80's to do so (i.e., Baumgardner and Rodi 1989). While Baumgardner and Rodi reported that independent laboratory calibration of the Rosemount icing probe improved its performance, they also reported that each of the four instruments tested had different mass sensitivities. There has been virtually no quantitative reporting of SLWC from an airborne Rosemount Icing probe in the literature. In time perhaps better SLWC measurement techniques applicable to TBS will be developed, implemented and reported.

References:

Baumgardner, D. and A. Rodi, 1989: Laboratory and Wind Tunnel Evaluations of the Rosemount Icing Detector. *J. Atmos. Oceanic Technol.*, **6**, 971–979.

Kitchen, M. and S. J. Caughey, 1981: Tethered-balloon observations of the structure of small cumulus clouds. *Q.J.R. Meteorol. Soc*., **107**, 853-874.

Lawson, R. P., K. Stamnes, J. Stamnes, P. Zmarzly, J. Koskuliks, C. Roden, Q. Mo, M. Carrithers, 2011: Deployment of a Tethered Balloon System for Cloud Microphysics and Radiative Measurements at Ny-Ålesund and South Pole, *J. Atmos. Oceanic Technol.* **28**, 656 – 670.

Lawson, R. P. and A. Gettelman. 2014: Impact of Antarctic mixed-phase clouds on climate. *Proc. Nat. Acad. Sci*., **111**, 18156–18161.